Computerized data-driven interpretation of the intrapartum cardiotocogram: a cohort study

Antoniya Georgieva*, PhD; Christopher W. G. Redman, MB, BChir, FRCP, FRCOG (ad eundum); Aris T. Papageorghiou, MBChB, FRCOG

Nuffield Department of Obstetrics & Gynaecology, University of Oxford, Women's Centre, John Radcliffe Hospital, Oxford OX3 9DU, UK.

*To whom correspondence should be addressed: tel.: +44 1865 857 854; email: antoniya.georgieva@obs-gyn.ox.ac.uk

Running headline: Computerized CTG analysis in labour

Conflicts of interest notification

None declared.

Abstract

Introduction: Continuous intrapartum fetal monitoring remains a significant clinical challenge. We propose utilising cohorts of routinely collected data. We aim to combine non-classical (data-driven) and classical cardiotocography (CTG) features with clinical features into a system (OxSys), which generates automated alarms for the fetus at risk of intrapartum hypoxia. We hypothesise that OxSys can outperform clinical diagnosis of 'fetal distress', when optimised and tested over large retrospective datasets.

Material and Methods: We studied a cohort of 22,790 labouring women (≥36 weeks gestation). Paired umbilical blood analyses were available. Perinatal outcomes were defined by objective criteria (Normal; Severe, Moderate or Mild compromise). We used the data retrospectively to develop a prototype of OxSys, by relating its alarms to perinatal outcome, and comparing its performance against standards achieved by bedside diagnosis.

Results:

OxSys1.5 triggers an alarm if the initial trace is nonreactive or the Decelerative Capacity (a non-classical CTG feature), exceeds a threshold, adjusted for preeclampsia and thick meconium. There were 187 newborns with Severe, 613 with Moderate and 3,197 with Mild compromise; and 18,793 with Normal outcome. OxSys1.5 increased the sensitivity for compromise detection: 43.3% vs. 38.0% for Severe ($p=0.3$) and 36.1% vs. 31.0% for Moderate ($p=0.06$); and reduced the false positive rate (14.4% vs. 16.3%, $p<0.001$).

Conclusions: Large historic cohorts can be utilised to develop and optimise computerized CTG monitoring, combining clinical and CTG risk factors. Our simple prototype has demonstrated the principle of using such data to trigger alarms, and compares well to clinical judgement.

Keywords: intrapartum fetal monitoring, computerized electronic fetal monitoring, CTG, sensitivity and specificity.

Abbreviations

CTG – cardiotocography

OxSys – Oxford System for computerized intrapartum fetal monitoring

FD^{clin} – Operative delivery in clinical practice for the clinical diagnosis of fetal distress

DC – Decelerative Capacity, a computerized feature of the CTG

Key Message

Large historic cohorts can be utilised to develop and optimise (prior to clinical testing) systems for data-driven computerized intrapartum monitoring that combine clinical and CTG risk factors.

Introduction

Cardiotocography (CTG) is widely used to continuously monitor the fetus during labour but its benefits are debated. Meta-analysis suggests that its routine use has a beneficial effect on the incidence of neonatal seizures (1). But it does not improve rates of cerebral palsy or perinatal mortality – while increasing instrumental and cesarean deliveries significantly (2). On the other hand, the technique has been credited by many with reducing the rate of intrapartum stillbirth: over the time that CTG was introduced intrapartum, stillbirth rates, in particular those associated with intrapartum hypoxia (3), have reduced. CTG is a possible cause for at least part of this improvement.

In this regard clinicians are trapped between two uncertainties. They are uncertain that CTG helps, but equally uncertain that it does not; and reluctant to discard a well-rehearsed technique with evidence from experience and basic research that it is effective, at least in part.

Important limitations of the current practice of visual and subjective fetal monitoring include a high false positive rate, which can lead to unnecessary cesarean or instrumental births (1,4); and a relatively low sensitivity, such that some babies at risk are not delivered in a timely fashion. Both are important, and shortcomings in labour management and CTG interpretation are leading causes for healthcare litigation (5). In USA, the litigation crisis has led some to suggest abandoning CTG altogether (6,7). Adjunct methods, such as fetal scalp blood sampling, fetal ECG analysis (such as STAN) and pulse oximetry have shown no consistent benefits (8-11).

The complex CTG patterns associated with fetal hypoxia are generally assessed visually; these have been shown to be poorly reproducible and inconsistent (12-14). However, CTG features can also be measured using computerized numerical analysis, which resolves this lack of reproducibility. Intrapartum, there are computerized systems that replicate “expert opinion” (15-19). In two randomised clinical trials, one showed no difference between computerized and subjective analysis (18); and the results from the other (INFANT (19)) are awaited. These systems merely emulate expert clinical assessment and have not been optimised on data to maximise the detection of adverse perinatal outcomes / minimise the false positive rate. In our opinion, computerization should go further: to use historic “big data” to discover new CTG features not apparent to the eye, measure relationship to outcome,

estimate the effect of confounding variables; ascertain the additive effect of clinical factors; and determine the background rates of CTG changes in the total population.

Based on this data-driven concept, we present here for the first time OxSys, an early prototype system demonstrating that non-classical and classical CTG features can be combined with clinical features to generate automated alarms for the fetus at risk of intrapartum hypoxia. The aims are, firstly, to demonstrate how cohort data can be used as a framework for development and evaluation of computerized CTG; and secondly to show how this enabled us to optimise the diagnostic accuracy of our early prototype system and test it in a way that is applicable to any similar CTG system.

Materials and Methods

Aim 1: Creating a framework for evaluation

The core composition of the Oxford archive has been described before (20,21) (Fig. 1). A total of 22,790ss births (dataset *Cord gases)* were included in this study. Singleton pregnancies were selected for the completeness of their CTG records acquired in routine clinical practice (both external and internal fetal monitoring was included); these constituted anonymized data from women in labour at 36 weeks of gestation or later, from March 2000 until December 2011 at the John Radcliffe Hospital, Oxford (Fig. 1). The study period ended in 2011 when our unit transferred to a new clinical data collection system. Pregnancies were excluded for problems that might affect the fetal responses to hypoxic stress, which would need separate scrutiny, such as prematurity, congenital malformations or breech delivery (Fig. 1).

Digitised intrapartum CTG data (sampled at 4 Hz) were archived by a central monitoring system. Clinical details were derived from the Oxford Clinical Maternity database and included basic maternal demographic and historical data, details of labour, delivery and perinatal outcome, including umbilical cord blood samples. The latter were considered to be valid if the difference between venous and arterial pH was at least 0.02 (22). The primary reason for cesarean or vaginal operative delivery at the time of the intevention (if applicable) was recorded electronically immediately after birth by the attending clinician. A drop down menu was used allowing eight possible reasons (fetal distress, failure to progress, prolonged second stage, placenta praevia, multiple birth, malpresentation, severe preeclampsia, previous obstetric history). This electronic record allowed us to distinguish a clinical diagnosis of fetal

distress (FD^{clin}) from other reasons for operative delivery but the precise time when the decision for operative delivery was taken, was not documented electronically. Only the records with validated cord gas analyses were included in this study (Fig. 1).

All CTG records, archived on our server were available for this retrospective analysis. For each woman, the entire CTG record was analysed. Those stopped three hours or more before delivery were excluded, because they could have little relevance to the analysis of features leading to obstetric intervention and/or affecting the condition of the baby at birth. Short fragments of CTG traces, less than 15 minutes long, were also excluded. Most traces (80%) ended less than five minutes before birth, with 93% ending less than 20 minutes and only 2% ending longer than 60min before birth.

Perinatal outcomes were classified into four exclusive groups (Table 1), defined pragmatically. Severe adverse outcome was defined as a composite outcome while . moderate and mild adverse outcomes were defined by the degree of acidemia in umbilical arterial blood.

We established when operative delivery was for FD^{clin} in each of the outcome groups. In parallel, retrospective occurrences of OxSys1.5 alarms were defined, which enabled us to compare OxSys to clinical management in terms of:

- The detection rates of different groups of adverse outcomes.
- The false positive rate, that is the comparative rates of positive tests associated with normal outcomes.
- A "dose-response" effect, i.e. whether an increasing number of alarms or interventions were observed in the outcomes with worsening compromise from Mild to Severe.

The framework we present is a method whereby computerized CTG analysis systems could be tested on a comprehensive database of unselected cases, which includes paired cord blood gas analyses at birth and, crucially, prospectively collected classification of the reasons for operative delivery during labour. The framework allows the likely utility of computerized CTG analysis systems to be evaluated; and allows calculation of effect size, sensitivity, false positive rates and potentially sample size to be estimated prior to prospective evaluation. In

this report we give the proof of principle of this method by evaluating our own computerized CTG system (an early prototype). This is described in Aim 2.

Aim 2: Applying the methodology to establish a prototype system for computerized intrapartum CTG analysis.

We assessed an early prototype computerized CTG system (The Oxford System, OxSys 1.5). The CTG is analysed in 15min windows; these move forward every five minutes, when the analysis is updated, on the basis of being a clinically relevant rate. This continues until an alarm is triggered or delivery, whichever occurs first. OxSys 1.5 is largely based on one computerized parameter – the decelerative capacity (DC) of Phase Rectified Signal Averaging (23-25). 'Decelerative capacity' analyses the entire fetal heart rate signal within the 15 minute analysis window, including accelerations and decelerations. It provides an average measure of downward movements in the fetal heart rate. Lower values are measured in a normal trace without significant decelerations. If the trace has many accelerations, these increase the DC but it remains well within the normal range (2bpm to 4bpm, data not shown). However, DC increases significantly if there are deep, steep-sloped and/or frequent decelerations. In effect, DC is a measure that combines both the time the trace spends at baseline and the frequency and depth of decelerations. Non-reactive traces have very low DC. DC has been confirmed to increase at the time of induced cord occlusion in the fetal lamb model of intrauterine hypoxic-ischaemic stress (24). The specific DC parameters were set as in our previous work (23). Based on an iterative process that uses both clinical knowledge and mathematical optimisation (26,27), we established the current system configuration (OxSys 1.5): a single OxSys alarm is triggered if: (1) the first hour of the trace is flat and non-reactive (DC value below 1bpm without accelerations, (28)); or (2) the DC reaches a defined threshold at any point during labour. The threshold is adjusted to a lower value if there is thick meconium or preeclampsia or to a higher value otherwise. A single OxSys alarm is required in order for us to count the alarm as a true or false positive.

Only CTG segments with valid signal in at least 50% of the time (signal quality) and only alarms that occurred 15 minutes or longer before the time of birth were considered. This is because alarms triggered nearer to the time of delivery would have been 'too late' to influence management.

The study was approved by the Newcastle & North Tyneside 1 Research Ethics Committee, REC reference 11/NE/0044 (data before 2008) and the South Central Ethics Committee, REC reference 13/SC/0153 (data beyond 2008). Informed consent was not required.

Results

The demographic characteristics of the women are reported in Table 2. The dataset *Cord gases* (included in our analysis in this study) had a higher rate of operative deliveries than the cohort from which it was derived, *Birth indication (*all monitored deliveries in March 2000-December 11). Also, slightly more babies showed thick meconium.

Dataset *Cord gases* includes 187 babies with Severe compromise. The clinical sensitivity or ability to detect Severe compromise during labour is 37.97% (Table 3). Thus, this project has enough power (α=0.05, β=0.1, two-sided Wald test) to detect significantly higher sensitivity for OxSys at 50% or above.

The size of the database is crucial for reliable estimation of the rate of OxSys alerts in Normals (false positive rate). A decrease in false positive rate of ≥0.7% can be detected in *Cord gases* (α=0.05, β=0.1, two-sided Wald test), compared to the estimated rate of clinical intervention in Normal – 16.33% (Table 3).

Optimisation and diagnostic accuracy of current prototype (OxSys1.5)

We utilised the above framework to iteratively develop a diagnostic system prototype (illustrated in Fig. 2): we began with a very simple prototype: one triggering an alarm if at any point the DC passed a single threshold (5.8bpm). After further experiments, we established that the sensitivity for severe compromise can be improved (without worsening the false positive rate) if the DC threshold was lowered to 4bpm in the presence of thick meconium or preeclampsia, which are well known clinical risk factors, and increased to 6.8bpm otherwise. These thresholds were selected after optimisation on the data. Furthermore, we established that the sensitivity to severe compromise was increased further if an alarm was also triggered in the rare cases where the initial CTG trace was nonreactive (Fig. 2).

The sensitivities of detecting Severe, Moderate or Mild adverse outcomes both by clinical assessment and OxSys1.5 are presented in Table 3. The detection rates for compromise types were consistently in favour of OxSys1.5 but the differences were statistically significant only

if Severe and Moderate compromise were combined into one category: 32.6% vs. 37.8%, Chi squared test, p=0.03.

The false positive rates can be measured in the Normal outcome group, in whom a clinical decision was taken to intervene (FD^{clin}) or the computerized system caused an "alarm". Here the respective figures were 16.3% for FD^{clin} and 14.5% for OxSys1.5, (p<0.001)

Thus OxSys1.5 is associated with comparable or higher sensitivity and significantly lower false positive rates when compared to clinical diagnosis. A dose-response relationship was demonstrated for both FD^{clin} and OxSys: there was an increasing rate of alarms from Normal through Mild and Moderate to Severe. Furthermore, the rates of both FD^{clin} and OxSys alarms declined with increasing pH threshold, in parallel, with OxSys alarm rates being consistently higher (Fig. 3), demonstrating a dose-response relationship.

An important aspect of intrapartum CTG is the degree of urgency of delivery when the monitoring is abnormal. Only carefully conducted prospective studies can show how useful the system is in this regard. We can retrospectively determine how long before birth an alarm would have been triggered by OxSys (the Alarm to Delivery (A-D) interval). This interval must be long enough to allow time for operative delivery. The minimal acceptable A-D interval depends on the clinical context, e.g. whether the woman is in the second stage of labour.

The results reported in Table 3 of our manuscript used an A-D of 15 minutes. However we also investigated varying A-D intervals (from 0 to 60min) and stratified the analysis according to the labour stage. In the first stage of labour, the performance of OxSys would have been similar if the A-D interval had been 60min. However, in the second stage of labour, the alarm rates quickly reduced as the A-D interval rose from 0min to 60min: only 45% of alarms in the second stage were raised 45min or more before birth, and 25% were raised 60min or earlier. Hence an alarm in the first stage of labour in most instances gave a reasonable warning time of at least 60 minutes; the shorter warning time in the second stage of labour is to be expected and is not necessarily incompatible with timely intervention.

Overall, regardless of labour stage, around 40% of OxSys alarms in babies with Severe compromise and about 20% in those with Moderate, were trigerred more than five hours prior to birth.

Discussion

We describe for the first time a methodology with which computerized CTG interpretation can be evaluated and compared to clinical assessment, using the same historic cohort data and identical measures of diagnostic accuracy. In this retrospective study, operative delivery was used to expedite delivery because of fetal distress (FD^{clin}) in 37.97% of cases with Severe adverse outcome and in 31.00% of those with Moderate. The rate of unnecessary intervention due to fetal distress was, as is generally believed, very high: over 16% interventions in the Normal group (a third of these were Caesareans and two thirds were vaginal operative deliveries). The simple prototype system presented here (OxSys1.5) already performed at least as well, if not slightly better, than clinical assessment (both in terms of higher sensitivity and lower false positive rate). In future work, we will adopt the framework and iterative development process illustrated here (Fig. 2), to define additional rules for triggering alarms. These will be added to OxSys to ensure a significantly better system prototype with substantially higher clinical utility.

The strengths of this study include our large cohort of high quality detailed data. We also considered various perinatal outcomes and the effects of different arterial pH thresholds (Fig. 3). Our data spans years and clinical practice has inevitably changed in that time. However, there was no evidence of any temporal change to the performance of OxSys during the study period (data not shown). Furthermore, a system that works well on a cohort spanning years of varying clinical practice, will be better positioned to work well on new data, than one designed with data from a narrow snapshot of clinical practice.

Retrospective data to evaluate methods for CTG have been previously used (18,34). The performance of PeriCALM (marketed by PeriGen), which simulates clinical expert assessment has been reported (34). However, the cases were selected based on being either severely compromised or with no signs of acidosis. Hence, a valid false positive rate is not available. In another study of >8000 deliveries the sensitivity of significant ST events from STAN monitoring (Neoventa Medical), (18) to predict arterial pH<7.05 was found to be 44.4%; this is higher than in our study (37.3% for OxSys) but at a higher alarm rate for all babies with arterial pH≥7.05 of 19.2% for STAN (16.1% for OxSys). It must be remembered that for STAN, accurate evaluation of CTG is also necessary, and combining STAN with OxSys is an interesting future possibility.

There are also inherent limitations and considerations, which are valid for any work with intrapartum CTG analysis:

Firstly, retrospective studies have inherent biases, and these can affect analysis of CTG data (30). A higher false positive rate and lower sensitivity is expected in populations with lower prevalence of compromise, but the proposed diagnostic methods are still valid, because the basic pathophysiological relationship between CTG patterns and compromise is universal.

Secondly, there is a 'treatment paradox': if perinatal compromise has been prevented by appropriate intervention, it could present as a false positive – a wider problem in obstetrics discussed recently (31). In this context, cord blood gas analyses and our definitions of outcome are important: it is reasonable to assume a false positive intervention in the absence of acidemia (the Normal group, (32)). However, if there is moderate acidemia, the outcome could be placed in a category of a prevented true positive (i.e. Moderate or Mild compromise).

Thirdly, the condition of the baby at birth cannot be known at the time the decision to deliver is made using CTG (neither in current clinical practice or with computerized alarms). The CTG is fundamentally limited in its ability to 'predict' intrapartum hypoxia – it is only a surrogate marker of measuremnets that we have no access to (for example fetal blood pressure, fetal cerebral perfusion, fetal oxygenation, etc.). However, in the absence of such direct measures, a 'risk assessment' approach by identifying a group of fetuses at increased risk of compromise, needs to be taken.

There are conflicting views as to whether acidemia alone is a valid endpoint of CTG and clinical obstetrics (30) prompting the need to define separate severe and moderate compromise groups. However, not all severe compromises have an intrapartum cause and are preventable by CTG monitoring: the fact that it cannot often be established whether a baby's injury was due to labour and/or preventable, is a wider challenge in clinical practice, litigation and CTG research. Therefore, we have kept an open mind about the aetiology of perinatal cerebral injury and proposed one overarching study design, by focusing on both clinically important severe compromise and different grades of academia. Our definitions of perinatal outcomes (Table 1) are based in part on umbilical arterial blood gas measurements at birth. Although there is no consensus as to how acidemia is best measured we have previously demonstrated that pH is preferable to Base Deficit (BD) (21). Moreover, in our study, 96% of the newborns with pH<7.05 had a BD>10mmol/l and 85% >12mmol/l. In Fig. 3, we demonstrate the effect of changes in pH thresholds, which amounts to a graded (and mechanistically reasonable) dose response effect.

Finally, CTG recordings during labour often may have poor quality leading to misinterpretation and uncertainty. This is a problem both in visual and computerized CTG interpretation. There are further limitations of the techniques for CTG acquisition, with rounding off errors and lack of 'true' beat to beat detailed data (as available in adult heart rate analysis).

All of the above limitations are inherent to CTG monitoring in clinical practice and our study does not claim to resolve these. Instead, our work proposes that further progress is possible despite these limitations and we propose to develop future OxSys versions with these considerations in mind.

At this stage, our early prototype has limited clinical utility. But we advocate the future development of refined systems using our framework to allow objective, standardised reporting and comparison between the diagnostic accuracies of different methods for CTG interpretation. Further work will examine the patterns of those fetuses with Severe compromise that were unrecognized and of those that triggered false alarms, generating new hypotheses about important CTG patterns. Any new hypotheses can be rapidly tested using the data leading to an iterative development of scientific, evidence-based methods for CTG interpretation. This means that computerized CTG can be extensively optimised before expensive clinical trials. In this regard validation on different independent datasets is essential prior to prospective clinical testing.

Funding

AG is funded by a generous research grant from Action Medical Research and the Henry Smith Charity.

References

1. Dyson Ch, Austin T. Lees Ch, Could routine cardiotocography reduce long term cognitive impairment? BMJ 2011;342:d3120.

2. Alfirevic Z, Devane D, Gyte GM. Continuous cardiotocography (CTG) as a form of electronic fetal monitoring (CTG) for fetal assessment during labour. Cochrane Database Syst Rev. 2013 May 31;5:CD006066.

3. Walsh CA, McMenamin MB, Foley ME, Daly SF, Robson MS, Geary MP. Trends in intrapartum fetal death, 1979-2003. Am J Obstet Gynecol. 2008 Jan;198(1):47.e1-7.

4. Vogel et al 2015. WHO Multi-Country Survey on Maternal and Newborn Health Research Network. Use of the Robson classification to assess caesarean section trends in 21 countries: a secondary analysis of two WHO multicountry surveys. Lancet Glob Health. 2015 May;3(5):e260-70.

5. The NHS Litigation Authority. Ten years of maternity claims: an analysis of NHS Litigation Authority data 2012.

6. Sartwelle TP. Electronic Fetal Monitoring: A Defense Lawyer's View. Rev Obstet Gynecol. 2012;5:e121-5.

7. Grimes DA, Peipert JF. Electronic fetal monitoring as a public screening program: the arithmetic failure. Obstet Gynecol. 2010;116(6):1297-400.

8. Olofsson P, Ayres-de-Campos D, Kessler J, Tendal B, Yli BM, Devoe L. A critical appraisal of the evidence for using cardiotocography plus ECG ST interval analysis for fetal surveillance in labour. Part I: the randomized controlled trials. Acta Obstet Gynecol Scand. 2014;93:556-8.

9. Saade G. Fetal ECG analysis of the ST segment as an adjunct to intrapartum fetal heart rate monitoring: a randomized clinical trial. Am J Obstet Gynecol. 2015;212:S2-S2 .

10. Chandraharan E. Fetal scalp blood sampling during labour: is it a useful diagnostic test or a historical test that no longer has a place in modern clinical obstetrics? *BJOG*. 2014;121:1056-60.

11. East CE, Begg L, Colditz PB, Lau R. Fetal pulse oximetry for fetal assessment in labour. *Cochrane Database Syst Rev* Oct 7;10:CD004075 2014.

12. Devane D, Lalor J. Midwives' visual interpretation of intrapartum cardiotocographs: intra- and inter-observer agreement. J Adv Nurs. 2005 Oct;52(2):133-41.

13. Rhöse S, Heinis AM, Vandenbussche F, van Drongelen J, van Dillen J. Inter- and intra-observer agreement of non-reassuring cardiotocography analysis and subsequent clinical management. Acta Obstet Gynecol Scand. 2014 Jun;93(6):596-602.

14. Chauhan SP, Klauser CK, Woodring TC, Sanderson M, Magann EF, Morrison JC. Intrapartum nonreassuring fetal heart rate tracing and prediction of adverse outcomes: interobserver variability. Am J Obstet Gynecol. 2008;199:623.e1-5.

15. Keith RDF, Beckley S, Garibaldi JM, Westgate J, Ifeachor EC, Greene KR. A multicentre comparative study of 17 experts and an Intelligent computer system for managing labour using the cardiotocogram. BJOG. 1995;102:688-700.

16. Elliot C, Warrick PA, Graham E, Hamilton EF. Graded classification of fetal heart rate tracings: association with neonatal metabolic acidosis and neurologic morbidity. Am J Obstet Gynecol. 2010;202:4258.e1-8.

17. Costa MA, Ayres-de-Campos D, Machado A, Santos CC, Bernardes J. Comparison of a computer system evaluation of intrapartum cardiotocographic events and a consensus of clinicians. J Perinat Med. 2010;38:191-5.

18. Nunes I, Ayres-de-Campos D, Ugwumadu A, Amin P, Banfield P, Nicoll A et al. for the FM-ALERT study group. FM-ALERT: a randomised clinical trial of intrapartum fetal monitoring with computer analysis and alerts versus previously available monitoring [abstract], ECIC'15, Porto, Portugal (2015).

19. INFANT clinical trial protocol [internet, cited 2015 Oct 07]. Available from http://www.ucl.ac.uk/cctu/researchareas/womenshealth/infant/information-for-clinicians.

20. Georgieva A, Payne SJ, Moulden M, Redman CW. Relation of fetal heart rate signals with unassignable baseline to poor neonatal state at birth. Med Biol Eng Comput. 2012 Jul;50(7):717-25.

21. Georgieva A, Moulden M, Redman CWG. Umbilical cord gases in relation to the neonatal condition: the EveREst plot. Eur J Obstet Gynaecol Reprod Biol. 2013;168:155-60.

22. Armstrong L, Stenson BJ. Use of umbilical cord blood gas analysis in the assessment of the newborn. Archives of Disease in Childhood Fetal and Neonatal 2007;92:F430–4.

23. Georgieva A, Papageorghiou AT, Payne SJ, Moulden M, Redman CWG. Phase rectified signal averaging for intrapartum electronic fetal heart rate monitoring is related to acidaemia at birth. BJOG. 2014;128:889-94.

24. Rivolta MW, Stampalija T, Casati Dm Richardson BS, Ross MG, Frasch MG, Bauer A, Ferrazzi E, Sassi R. Acceleration and deceleration capacity of fetal heart rate in an in-vivo sheep model. PLoS One 2014;9(8):e104193.

25. Advances in computing are driving progress in fetal monitoring, A. Georgieva. British Journal of Obstetrics and Gynaecology Dec 25, 2015 DOI: 10.1111/1471-0528.13725.

26. Optimization in Medicine, eds. CJS Alves, PM Pardalos, LN Vicente, Springer 2008;

27. Optimization in Medicine and Biology, eds. GJ Lim and EK Lee, Auerbach Publications 2008

28. Georgieva A, Ugwumadu A, Papageorghiou AT, Redman CWG. Significance of the first hour of the fetal heart rate monitoring: nonreactive vs reactive initial trace, *European Congress on Perinatal Medicine*, Maastricht, 2016.

29. Yudkin PL, Aboualfa M, Eyre JA, Redman CW, Wilkinson AR. New birthweight and head circumference centiles for gestational ages 24 to 42 weeks. Early Hum Dev 1987; 15(1): 45–52.

30. Cahill AG and Macones GA. Fetal heart rate tracings and neonatal metabolic acidosis: Elliott et al, Am J Obstet Gynecol. 2010;202:317-18.

31. F. Cheong-See, J. Allotey, N. Marlin, B.W. Mol, E. Schuit, G. ter Riet, R.D. Riley, K.G.M. Moons, K.S. Khan and S. Thangaratinam. Prediction models in obstetrics: understanding the treatment paradox and potential solutions to the threat it poses. *BJOG* published online: 25 Jan 2016 DOI: 10.1111/1471-0528.13859.

32. ACOG Task Force on Neonatal Encephalopathy, Neonatal encephalopathy and neurologic outcome, second edition (2014).

33. A. Bhide, E. Chandraharan, and G. Acharya. Fetal monitoring in labor: Implications of evidence generated by new systematic review. *Acta Obstet Gynecol Scand* 95, 5-8 (2015).

34. Elliot C, Warrick PA, Graham E, Hamilton EF. Graded classification of fetal heart rate tracings: association with neonatal metabolic acidosis and neurologic morbidity. Am J Obstet Gynecol. 2010;202:4258.e1-8.

415

416

417 **Tables**

418 Table 1. Definitions of the outcome groups and incidence in the dataset included in analysis.

Exclusive outcome groups		**Data included in this study: dataset *Cord gases*, n = 22,790**
Severe	Composite outcome of: stillbirth; neonatal death; neonatal encephalopathy; intubation or cardiac massage followed by admission to neonatal intensive care for ≥48hrs.	187 (0.82%)
Moderate	Umbilical cord arterial pH<7.05 without severe compromise	613 (2.69%)
Mild	7.05 ≤ pH < 7.15 without severe compromise	3,197 (14.03%)
Normal	pH >= 7.15 without severe compromise	18,793 (82.46%)

419

420

421 Table 2. The Oxford datasets: clinical and demographic characteristics: *n* (%). The datasets
422 are defined in Fig. 1.

	All (58,748)	*Birth indication* (38,818)	*Cord gases* (22,790)
Nulliparous	32,523 (55.36%)	21,927 (56.49%)	13,948 (61.20%)
Pre-eclampsia	6,303 (10.73%)	3,497 (9.00%)	2,282 (10.01%)
Gestational Diabetes	509 (0.87%)	423 (1.09%)	262 (1.15%)
Induction of labour	25,157 (42.82%)	17,966 (46.28%)	10,838 (47.56%)
Cesarean or Forceps/Ventouse delivery	22,432 (38.18%)	15,594 (40.17%)	11,382 (49.94%)
Cesarean	7,080(12.05%)	5,164 (13.30%)	3,908 (17.15%)
Thick meconium	4,531 (7.71%)	3,437 (8.85%)	2,474 (10.86%)
Oxytocin	22,062 (37.55%)	15,295 (39.40%)	9,783 (42.93%)
Low Apgar	1,448 (2.47%)	1,008 (2.60%)	782 (3.43%)
Severe compromise	473 (0.81%)	255 (0.66%)	167 (0.79%)
Convulsions	109 (0.18%)	74 (0.19%)	50 (0.22%)
Neonatal encephalopathy	164 (0.28%)	80 (0.20%)	51 (0.22%)
Intubation or cardiac massage	614 (1.05%)	318 (0.82%)	247 (1.08%)
SCBU admission ≥ 48hrs	2,910 (3.25%)	1,215 (3.13%)	811 (3.56%)
Small baby (<3rd centile§)	936 (1.59%)	575 (1.48%)	346 (1.52%)
Large baby (>97th centile§)	2,922 (4.97%)	1,953 (5.03%)	1,282 (5.64%)

Stillbirth	2 (0.0034%)	1 (0.0026%)	0 (0.00%)
Neonatal death (<28 days)	30 (0.05%)	22 (0.06%)	17 (0.07ss%)

§ adjusted Yudkin (29) centile.

Table 3. OxSys1.5 alarm rate and emergency deliveries in clinical practice due to fetal distress (FD^{clin}): number (%) [95% confidence interval].

Diagnostic accuracy on *Cord gases* dataset (*n*=22,790)				
Exclusive outcome groups	**Compromise** (*sensitivity*)			**Normal** (*false positive rate*)
	Severe	**Moderate**	**Mild**	
Number of births	187	613	3,197	18,793
Detected in clinical practice (FD^{clin})	71 (37.97%) [31.0%-44.9%]	190 (31.00%) [27.3%-34.7%]	719 (22.49%) [21.0%-23.9%]	3,068 (16.33%)* [15.8%-16.9%]
Detected by OxSys 1.5	81 (43.32%)** [36.2%-50.4%]	221 (36.05%)*** [32.2%-39.9%]	789 (24.68%)† [23.2%-26.8%]	2,710 (14.42%)‡ [13.9%-14.9%]

* 906 (29.5%) of these were Cesarean sections;

Chi squared test, OxSys 1.5 vs. Clinicians: ** $p = 0.29$; *** $p = 0.06$; †$p < 0.04$; ‡ $p<0.001$.

Figures

Fig. 1. Data flow chart. – in separate pdf.

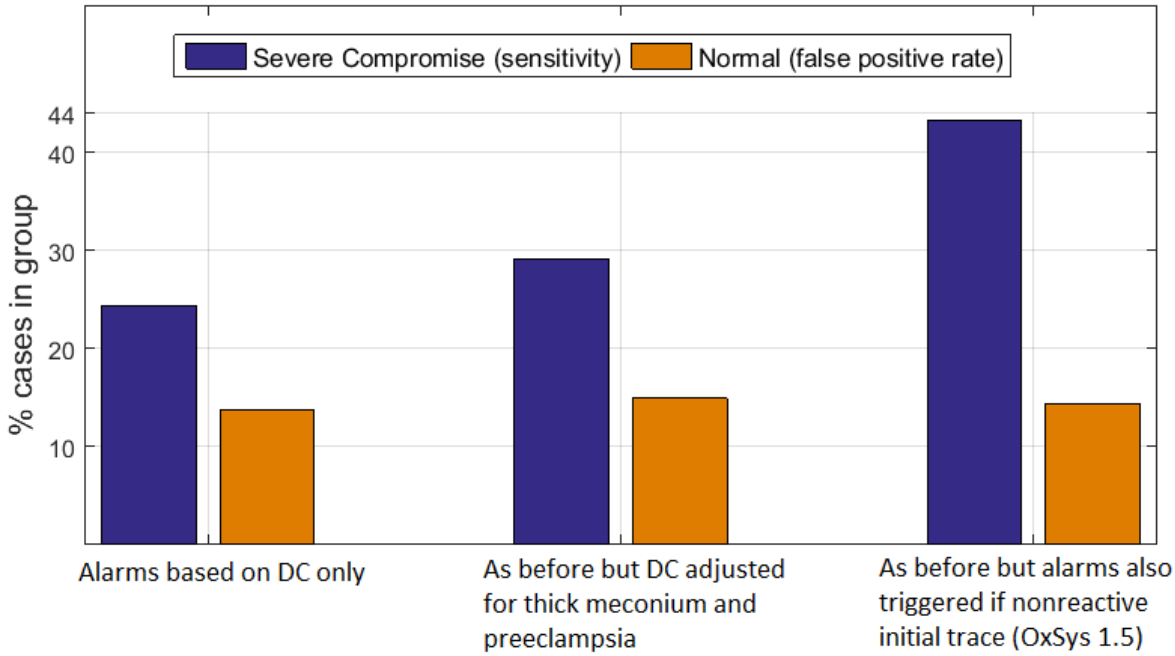

Fig. 2. OxSys evolution throught the iteration phases: the sensitivity of the system improved significantly ($p<0.001$) without increasing the false positive rate, *Cord gases* dataset (n = 22,790).

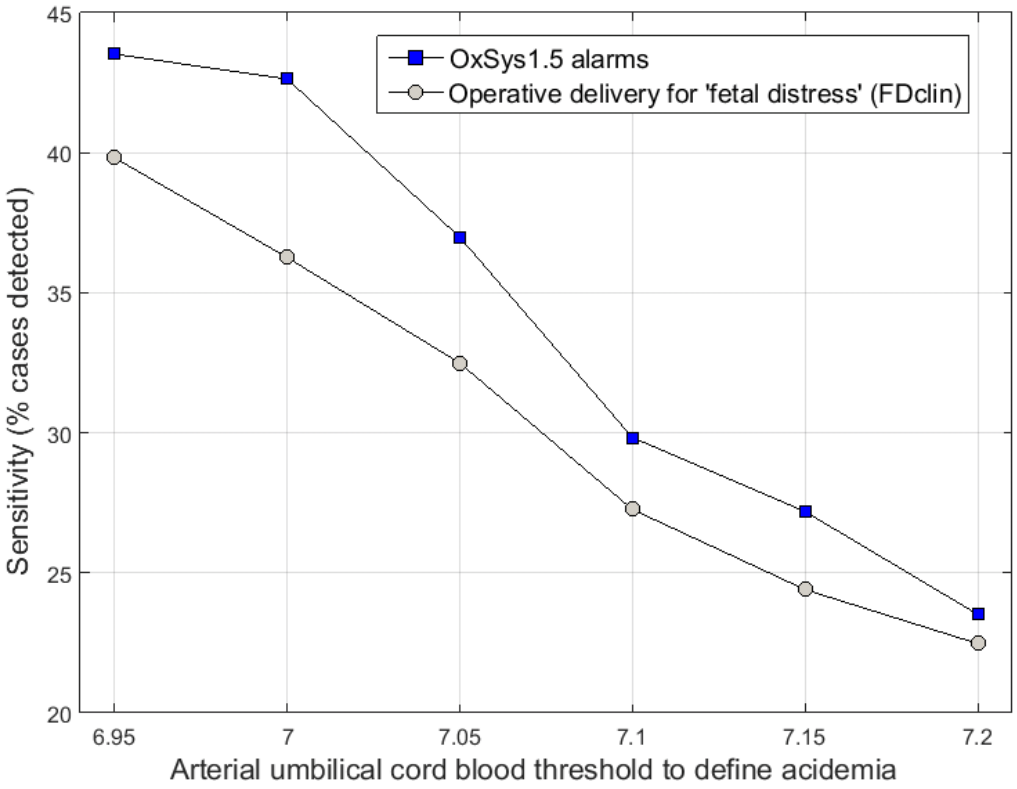

Fig. 3. The rate of emergency interventions for fetal distress (FD^{clin}) and the OxSys1.5 alarm rate (i.e. the sensitivity) consistently increase if lower pH threshold is used to define increasing acidemia, *Cord gases* dataset (n = 22,790).